Credibilistic Fuzzy Clustering Method Based on Evolutionary Approach of Crazy Wolfs in Online Mode

Alina Shafronenko\textsuperscript{a}, Yevgeniy Bodyanskiy\textsuperscript{a}, Iryna Pliss\textsuperscript{a}

\textsuperscript{a} Kharkiv National University of Radio Electronics, Nauky ave 14, Kharkiv, 61166, Ukraine

Abstract
The problem of big data credibilistic fuzzy clustering is considered. This task was interested, when data are fed in both batch and online modes and has a lot of the global extremums. To find the global extremum of the objective function of plausible fuzzy clustering, a modification of the crazy gray wolfs algorithm was introduced, which combines the advantages of evolutionary algorithms and global random search. It is shown that different search modes are generated by a unified mathematical procedure, special cases of which are well-known algorithms for both local and global optimization. The proposed approach is quite simple in computational implementation and is characterized by high speed and reliability in tasks of multi-extreme fuzzy clustering.

Keywords
Fuzzy clustering, credibilistic fuzzy clustering, adaptive goal function, evolutionary algorithms, gray wolf optimization

1. Introduction
The task of classification in the mode of self-learning (clusterization) of multidimensional data is an important part of traditional intellectual analysis.

One of the main directions of computational intelligence is the so-called evolutionary algorithms, which are mathematical models of the evolution of biological organisms. The problem of clustering associated with vector observations often arises in many tasks of intelligent data analysis, such as Data Mining, Dynamic Data Mining, Data Stream Mining, Big Data Mining, Web Mining [1, 2] and primarily in fuzzy data clustering, when the processing of vector observations with different levels of probability, reliability, etc. can belong to more than one class.

The most effective are Kohonen's self-organizing maps [3] and evolutionary algorithms, which can improve data clustering in cases where data is processed in online modes.

From a computational point of view, the problem of clustering is reduced to finding the optimum of multi-extremal functions of a vector argument with the help of gradient procedures, which are repeatedly launched from different starting points. For speed up the process of finding extremums by using the ideas of evolutionary optimization.

Evolutionary computation [4] and swarm intelligence are considered the fastest growing algorithms that employ computational intelligence to solve optimization problems. The evolutionary swarm intelligence includes genetic algorithm [5], biogeography-based optimizer [6], evolutionary strategies [7], population-based incremental learning [8], genetic programming [9], and differential evolution [10].

Swarm intelligence [11] is another powerful form of the computational intelligence used to solve the optimization problems. Swarm intelligence algorithm using the ideas of evolutionary optimization, which includes algorithms inspired by nature, swarm algorithms, population algorithms,
and imitate the natural swarms or communities or systems such as fish schools, bird swarms, cat swarms, bacterial growth, insect’s colonies and animal herds etc. Evolutionary swarm intelligence algorithms include many algorithms such ant colony optimization [12], particle swarm optimization (PSO) [13], artificial bee colony [14], ant lion optimizer (ALO) [15], moth-flame optimization, dragonfly algorithm, sine cosine algorithm, whale optimization algorithm, multi-verse optimizer, grey wolf optimizer (GWO) [16], firefly algorithm, cuckoo search, and many others.

The purpose of the paper is to consider the problem of data fuzzy clustering that is received for processing in both batch and online modes. To find the global optimum of a multi-extremal objective function, to develop a method that can find the global extremum of complex functions.

2. Problem Statement

Baseline information for solving the tasks of clustering in a batch mode is the sample of observations, formed from \(N\) \(n\)-dimensional feature vectors \(X = \{x_1, x_2, ..., x_N\} \subset \mathbb{R}^n, x_i \in X, k = 1, 2, ..., N\).

The problem of fuzzy clustering of data arrays is considered in the conditions when the formed clusters arbitrarily overlap in the space of features. The source information for solving the problem is an array of multidimensional data vectors, formed by a set of vector-observations \(X = (x(1), x(2), ..., x(k), ..., x(N)) \subset \mathbb{R}^n\) where \(k\) in the general case the observation number in the initial array, \(x(k) = (x_i(k), ..., x_{n}(k) x_{n}(n))^T\). The result of clustering is the partition of this array on \(m\) overlapped classes with centroids \(c_q \in \mathbb{R}^n, q = 1, 2, ..., m\), and computing of membership levels \(0 \leq u_q(k) \leq 1\) of each vector-observation \(x(k)\) to every cluster \(c_q\).

3. Credibilistic Fuzzy Clustering Method

Alternatively, to probabilistic and possibilistic procedures [17,18] it was introduced credibilistic fuzzy clustering approach using as its basis the credibility theory [19, 20] and is largely devoid of the drawbacks of known methods.

The most common approach within the framework of probabilistic fuzzy clustering is associated with minimizing the goal function [3].

\[
\text{Goal} \left( u_q(k), c_q \right) = \sum_{k=1}^{N} \sum_{q=1}^{m} u_q^\beta(k) d^2 \left( x(k), c_q \right) 
\]

where \(u_q(k)\) - membership level, \(c_q\) - centroid, \(d\) - Euclidian metric, \(\beta\) - fuzzifier;

with constraints

\[
\begin{align*}
\sum_{q=1}^{m} u_q(k) &= 1, \\
0 &< \sum_{k=1}^{N} u_q(k) < N.
\end{align*}
\]

Solution of nonlinear programming problem using the method of Lagrange indefinite multipliers leads to the well-known result
\[
\begin{align*}
\begin{cases}
    u_q(k) &= \left( \frac{d^2(x(k),c_q)}{\sum_{j=1}^{n} d^2(x(k),c_j)} \right)^{\frac{1}{\beta}}, \\
    c_q &= \frac{\sum_{k=1}^{N} u_q(k)^\beta x(k)}{\sum_{k=1}^{N} (u_q(k))^\beta},
\end{cases}
\end{align*}
\]

(3)

coinciding with \( \beta = 2 \) with a popular method of Fuzzy C-Means of J. Bezdek (FCM) [21].

If the data are fed to processing sequentially, the solution of the nonlinear programming problem (1), (2) using the Arrow-Hurwitz-Uzawa algorithm leads to an online procedure:

\[
\begin{align*}
\begin{cases}
    u_q(k+1) &= \left( \frac{d^2(x(k+1),c_q(k))}{\sum_{j=1}^{n} d^2(x(k+1),c_j)} \right)^{\frac{1}{\beta}}, \\
    c_q(k+1) &= c(k) + \eta(k+1)u_q^\beta(k+1),
\end{cases}
\end{align*}
\]

(4)

\[
* \left( x(k+1) - c_q(k) \right).
\]

The goal function of credibilistic fuzzy clustering has the form [19, 22] close to (1)

\[
Goal(Cred_q(k),c_q) = \sum_{k=1}^{N} \sum_{q=1}^{Q} Cred^\beta_q(k)d^2(x(k),c_q)
\]

(5)

with "softer" than (2) constraints:

\[
\begin{align*}
    0 \leq Cred_q(k) &\leq 1, \text{ for all } q \text{ and } k, \\
    \sup Cred_q(k) &\geq 0.5, \text{ for all } k, \\
    Cred_q(k) + \sup Cred_l(k) &= 1, \\
    \text{for any } q \text{ and } k, \text{for which } Cred_q(k) \geq 0.5.
\end{align*}
\]

(6)

It should be noted that the goal functions (1) and (5) are similar and that there are no rigid probabilistic constraints in (6) on the sum of the membership in (2).

In the procedures of credibilistic clustering, there is also the concept of fuzzy membership, which is calculated using the neighborhood function of the form

\[
u_q(k) = \varphi_q \left( d \left( x(k), c_q \right) \right)
\]

(7)

monotonically decreasing on the interval \([0, \infty]\) so that \( \varphi_q(0) = 1, \varphi_q(\infty) \to 0 \).

Such a function is essentially an empirical similarity measure of [23] related to distance by the relation

\[
u_q(k) = \frac{1}{1 + d^2(x(k),c_q)}.
\]

(8)

Note also that earlier it was shown in [16] that the first relation (3) for \( \beta = 2 \) can be rewritten as

\[
u_q(k) = \left( 1 + \frac{d^2(x(k),c_q)}{\sigma^2_q} \right)^{-1},
\]

(9)

where

\[
\sigma^2_q = \sum_{j=1}^{n} d^2(x(k),c_j)^{-1}
\]

(10)
which is a generalization of the function (8) (for $\sigma_q^2 = 1$ (8) coincides with (10)) and satisfies all the conditions for (7).

In batch form the algorithm of credibilistic fuzzy clustering in the accepted notation can be written as [20, 22]

$$
\begin{align*}
  u_q(k) &= \left(1 + d^2(x(k), c_q)\right)^{-1}, \\
  u^*_q(k) &= u_q(k)\left(\sup_{i \neq q} u_i(k)\right)^{-1}, \\
  \text{Cred}_q(k) &= \frac{1}{2}\left(u^*_q(k) + 1 - \sup_{i \neq q} u^*_i(k)\right), \\
  c_q &= \frac{\sum_{k=1}^{N} \left(\text{Cred}_q(k)\right)^{\beta} x(k)}{\sum_{k=1}^{N} \left(\text{Cred}_q(k)\right)^{\beta}}
\end{align*}
$$

and in the online mode, taking into account (9), (10):

$$
\begin{align*}
  \sigma^2_q(k + 1) &= \frac{1}{m} \sum_{i \neq q} d^2(x(k + 1), c_i(k)), \\
  u_q(k + 1) &= \left(1 + \frac{d^2(x(k + 1), c_q(k))}{\sigma^2_q(k + 1)}\right)^{-1}, \\
  u^*_q(k + 1) &= \frac{u_q(k + 1)}{\sup_{i \neq q} u_i(k + 1)}, \\
  \text{Cred}_q(k + 1) &= \frac{1}{2}\left(u^*_q(k + 1) + 1 - \sup_{i \neq q} u^*_i(k)\right), \\
  c_q(k + 1) &= c_q(k) + \eta(k + 1)\text{Cred}^\beta_q(k + 1)\left(x(k + 1) - c_q(k)\right).
\end{align*}
$$

From the point of view of computational implementation, algorithm (12) is not more complicated than procedure (4) and, in the general case, is its generalization to the case of credibilistic approach to fuzzy clustering.

4. **Evolutionary Approach of Crazy Wolfs**

To find the global extremum of functions (1) and (5), it is advisable to use bio-inspired evolutionary algorithms for particle swarm optimization [24], among which the so-called gray wolf algorithm (GWO) can be noted as one of the fastest coding ones [25]. According to the algorithm proposed in [25], gray wolves live together and hunt in groups.

The search and hunting process can be described as follows:

$$
\begin{align*}
  C_a &= |B_1 \ast GW^a - X(t)|, \\
  C_b &= |B_2 \ast GW^b - X(t)|, \\
  C_\beta &= |B_3 \ast GW^\beta - X(t)|,
\end{align*}
$$

if the prey is found, they first track, pursue and approach it.

If the prey runs, then the gray wolves chase, surround and watch the prey until it stops moving, which can be written as
\[ GW_1 = GW_a - A_1 \cdot C_a, \]
\[ GW_2 = GW_{\beta} - A_2 \cdot C_{\beta}, \]
\[ GW_3 = GW_{\delta} - A_3 \cdot C_{\delta}, \]

where \( GW \) - wolf movement function.

After the prey is found, the attack begins:
\[ GW(t) = \frac{GW_1 + GW_2 + GW_3}{3}. \]

**Figure 1:** How \( \alpha-, \beta-, \delta- \) - wolves are defined in GWO

In a case, when \(|A| > 1\) - gray wolves flee from dominants, which means that omega wolves will flee from prey and explore more space; if \(|A| < 1\) they approach the dominants, which means that \( \delta\) - wolves will follow the dominants that approach the prey.

**Figure 2:** Exploration and exploitation of wolf in GWO
To achieve a proper trade-off between scouting and hunting, an improved gray wolf algorithm is proposed.

**Figure 3:** Scheme of operation of the gray wolf algorithm

The change in the positions of the wolves is described as follows:

\[
C = |B^*X_P(t) - X(t)|, \tag{10}
\]

\[
X(t + 1) = X_P(t) - A^*C + \Xi(\tau), \tag{11}
\]

where \(X_P\) - prey position, \(X\) - position of the wolf, \(\Xi(\tau)\) - a random perturbation that introduces an additional scan of the search space.

To increase the reliability of finding exactly the global extremum of the objective function, you can use the idea of "crazy cats" - optimization [26, 27], modified by the introduction of a random walk, which has proven its effectiveness when solving multi-extremal problems. By introducing, according to L. Rastrigin, an additional search perturbation, it is possible to write down the movement of the wolf in the form:

\[
\Xi(\tau) = \gamma\Xi(\tau - 1) - \delta(X(t + 1) - X(t)) + \sigma^2 H(\tau) \tag{12}
\]

where \(\gamma\) - parameter for correction of disturbance characteristics, \(0 \leq \delta \leq 1\) - self-learning speed parameter, \(\sigma^2\) - white noise variance \(H(\tau)\).

In this way, the probability of finding the global extremum of the adopted objective function increases, which ultimately increases the effectiveness of the fuzzy clustering process.

## 5. Experimental Research

The proposed method was tested on well-known multiextremal function such as Ackley's function, presented in form (13), Fig. 4:

\[
f(x) = -20\exp(-0.2\sqrt{0.5(x^2 + y^2)}) - \exp(0.5\cos(2\pi y)) + e + 20. \tag{13}
\]
Figure 4: Ackley's function

Pseudocode of proposed method can write in the form of next steps presented in the Table 1:

<table>
<thead>
<tr>
<th>Description</th>
<th>Pseudocode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup options</td>
<td>Data sampling</td>
</tr>
<tr>
<td></td>
<td>Swarm population</td>
</tr>
<tr>
<td></td>
<td>Control parameter</td>
</tr>
<tr>
<td></td>
<td>Stop criterion</td>
</tr>
<tr>
<td>Initialization</td>
<td>Initial positions of dominant gray wolves</td>
</tr>
<tr>
<td>Search</td>
<td>If this is not a stopping criterion, then the calculation of the new value of the fitness function</td>
</tr>
<tr>
<td></td>
<td>Position update</td>
</tr>
<tr>
<td></td>
<td>Restrictions on the positions of wolves</td>
</tr>
<tr>
<td></td>
<td>Update $\alpha$, $\beta$ and $\delta$</td>
</tr>
<tr>
<td></td>
<td>Update stop criteria</td>
</tr>
<tr>
<td></td>
<td>End</td>
</tr>
</tbody>
</table>

The behavior of Crazy GWO method can tested on multiextremal Ackley's function. Analyzes was provide on 100 iterations. On Fig. 5 demonstrate search history of global extremum of proposed function (a), the trajectory of the first variable of the first wolf (b), fitness history of all wolves (c) and the parameter $A$ (d).
For further validating the searching accuracy of Crazy GWO method (CGWO) with GWO and PSO algorithms.

**Table 2**

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>CGWO</th>
<th>GWO</th>
<th>PSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>1.5099e-017</td>
<td>7.5495e-013</td>
<td>0.0035</td>
</tr>
<tr>
<td>Ave</td>
<td>1.2204e-016</td>
<td>1.0048e-012</td>
<td>0.0055</td>
</tr>
<tr>
<td>Worst</td>
<td>2.2204e-014</td>
<td>1.4655e-013</td>
<td>0.0082</td>
</tr>
</tbody>
</table>

Two swarm intelligence algorithms (PSO algorithm and GWO algorithm) are selected to carry out the simulation contrast experiments with the proposed CGWO so as to verify its superiority on the convergence velocity and searching precision. The simulation convergence results on the adopted testing functions are shown in Fig. 6:
6. Discussion

Analyzing the results of the obtained experimental studies and comparative analysis of the method of clustering data arrays based on the crazy gray wolf algorithm with clustering methods based on both the classical approach to data clustering and more exotic ones, the proposed method demonstrates sufficiently high results.

The main advantages of the proposed method are the simplicity of mathematical calculations, the speed of working with data, regardless of the type, size, and quality of the analyzed sample. It should be noted the accuracy of the data clustering method based on the improved gray algorithm and the obtained clustering results, which is achieved with the help of the optimization procedure of the evolutionary algorithm.

Conclusion

The problem of credibilistic fuzzy clustering of data using a crazy gray wolf algorithm is considered. The proposed method excludes the possibility of "getting stuck" in local extrema by means of a double check of finding the dominant wolf in the extremum and, in comparison with the given calculation error, allows to reduce the number of procedures starts. A feature of the proposed modified optimization procedure is the possibility of managing random disturbances that determine the properties of the modified the random walk, that has proven effectiveness when solving multi-extremal problems.

The approach is quite simple in numerical implementation, allows finding global extrema of complex functions, which is confirmed by the results of experimental research.

7. Acknowledgement

The work is supported by the state budget scientific research project of Kharkiv National University of Radio Electronics “Development of methods and algorithms of combined learning of deep neuro-neo-fuzzy systems under the conditions of a short training sample”.

8. References